Empirical Database

Víctor Javier España Roch

17/12/2020

library(eat)  
setwd("C:/Users/victo/Desktop/UNIVERSIDAD/EAT/simulaciones")  
library(readxl)  
airline <- read\_excel("airline.xlsx")

## Airline database

Random Forest EAT (RFEAT) has also been developed with the aim of providing a greater stability to the results obtained by EAT. The function requires the data containing the variables for the analysis, x and y corresponding to the inputs and outputs indexes respectively, the minimun number of observation in a node for a split to be attempted (numStop) and na.rm to ignore observations with NA cells. All these arguments are used for the construction of the m individual EAT trees that make up the random forest. Finally, the argument s\_mtry indicates the number of inputs that can be selected in each split. Being, the number of inputs, the number of outputs and the sample size, the available options in s\_mtry are:

* Breiman =
* DEA1 =
* DEA2 =
* DEA3 =
* DEA4 =

RFEAT\_airline <- RFEAT(data = airline,  
 x = 1:2,  
 y = 3:6,  
 numStop = 1,  
 m = 50,  
 s\_mtry = "Breiman",  
 na.rm = TRUE)

Predictions for the RFEAT model are made using the predict\_RFEAT function. To do this, the output is predicted by each of the m individual trees trained and subsequently the mean value of all the predictions is obtained. predict\_RFEAT requires a RFEAT object (from RFEAT function) and a set of input variables to predict on (newdata) returning then, a dataframe with the inputs and predicted variables.

predictions\_RFEAT <- predict\_RFEAT(object = RFEAT\_airline,  
 newdata = airline[, 1:2])

## y1\_pred y2\_pred y3\_pred y4\_pred  
## 1 46.22 14.76 9.72 5.120  
## 2 48.16 14.88 8.94 5.490  
## 3 52.34 20.56 13.52 6.160  
## 4 37.64 10.86 5.42 2.394  
## 5 57.30 18.92 13.52 6.160  
## 6 33.68 10.86 3.96 1.640  
## 7 31.26 9.02 4.04 1.556  
## 8 40.34 14.76 7.18 3.170  
## 9 37.54 10.86 5.42 2.294  
## 10 50.32 15.90 13.18 5.540

Finally, a specific method to calculate efficiency scores applied to RFEAT can be utilized using the efficiency\_RFEAT function.

scoresRF <- efficiency\_RFEAT(data = airline,  
 x = 1:2,  
 y = 3:6,  
 object = RFEAT\_airline)

## scoreRF  
## 1 0.883636  
## 2 0.784286  
## 3 0.893913  
## 4 1.330000  
## 5 0.955000  
## 6 0.905000  
## 7 1.127500  
## 8 0.868235  
## 9 0.764667  
## 10 0.878667